COMBINE Analysis of the Specificity of Binding of Ras Proteins to their Effectors

Sanja Tomić,1,2 Branimir Bertoša,1 Ting Wang,3 and Rebecca C. Wade4
1Ruđer Bošković Institute, Bijenička 54, Zagreb, Croatia
2Molecular and Cellular Modeling Group, EML Research gGmbH, Schloss-Wolfsbrunnenweg 33, Heidelberg, Germany

ABSTRACT The small guanosine triphosphate (GTP)-binding proteins of the Ras family are involved in many cellular pathways leading to cell growth, differentiation, and apoptosis. Understanding the interaction of Ras with other proteins is of importance not only for studying signalling mechanisms but also, because of their medical relevance as targets, for anticancer therapy. To study their specificity and activity, which are essential to their signal transduction function, we performed COMparative Binding Energy (COMBINE) analysis for 122 different wild-type and mutant complexes between the Ras proteins, Ras and Rap, and their effectors, Rap and Rap1GDS. The COMBINE models highlighted the amino acid residues responsible for subtle differences in binding of the same effector to the two different Ras proteins, as well as more significant differences in the binding of the two different effectors (RapGDS and Rap) to Ras. The study revealed that E37, D38, and D57 in Ras are nonspecific hot spots at its effector interface, important for stabilization of both the RapGDS-Ras and Rap-Ras complexes. The electrostatic interaction between a GTP analogue and the effector, either Rap or Rap1GDS, also stabilizes these complexes. The Rap-Ras complexes are specifically stabilized by S38, Y40, and D54, and RapGDS-Ras complexes by E31 and D33. Binding of a small molecule in the vicinity of one of these groups of amino acid residues could increase discrimination between the Rap-Ras and RapGDS-Ras complexes. Despite the different size of the RapGDS-Ras and Rap-Ras complexes, we succeeded in building COMBINE models for one type of complex that were also predictive for the other type of protein complex. Further, using system-specific models trained with only five complexes selected according to the results of principal component analysis, we were able to predict binding affinities for the other mutants of the particular Ras-effector complex. As the COMBINE analysis method is able to explicitly reveal the amino acid residues that have most influence on binding affinity, it is a valuable aid for protein design. Proteins 2007;67:000-000. © 2007 Wiley-Liss, Inc.

Key words: Ras; RapGDS; Rap; Rap1GDS; binding affinity; binding specificity; Poisson-Boltzmann electrostatics; hot spots; quantitative structure-activity relationship

INTRODUCTION Signal transfer in biological systems depends on protein-protein interactions. Protein mutations and changes in environment can abolish signalling pathways within cells, leading to severe damage and disease. The first identified and the most studied role of Ras family members is in cell transformation. The small guanosine triphosphate (GTP)-binding proteins of the Ras family are involved in many cellular pathways leading to cell growth, differentiation, and apoptosis.1 Regulatory proteins, guanine nucleotide exchange factors (GEFs), and GTPase activating proteins (GAPs) turn Ras, and Ras-related GTP-binding proteins, on and off.2 In the active, on conformation, Ras interacts with the effector Rap, a Ser/Thr specific protein kinase that is an immediate downstream target of Ras in the mitogen-activating protein kinase pathway (MAPK).1,3 Certain mutations in Ras destroy this switch function. In the mutated Ras protein, the hydrolysis of GTP to GDP is not possible and Ras remains in its active on conformation. In this way, the signal transduction is amplified and cells grow and divide uncontrolled. The mutated Ras, as well as other proteins involved in Ras signal transduction, are frequently found in diverse human tumors.2,5 Mutations in

Abbreviations: COMBINE, COMparative BINDing Energy; GDP, guanosine diphosphate; GEF, guanine diphosphate exchange factor; GTP, guanosine triphosphate; GC, principal components; PBS, partial least squares; QSAR, quantitative structure activity relationship; Rap; Rap1GDS; RapGDS; guanine nucleotide dissociation inhibitor--stimulator of the small GTPase Rap; Rap1A—member of the Ras small G protein superfamily; Rap; Rap protein product of human Hn-Ras; RBD, Ras binding domain of effector molecules; RMSD, root mean square deviation.

*Correspondence to: Sanja Tomić, Ruđer Bošković Institute, Bijenička 54, Zagreb, Croatia. E-mail: sanja.tomic@rb.hr
Received 4 August 2006; Revised 10 October 2006; Accepted 6 November 2006
Published online 00 Month 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/pro.21321

© 2007 WILEY-LISS, INC.
Ras that permanently activate it have been found in about half of all colon cancers and over 95% of pancreatic cancers. Another effector protein that Ras–proteins interact with is Raf guanine nucleotide dissociation stimulator (RafGDS). This interaction regulates Rho/Rac family members that control cytoskeletal rearrangements. There is also experimental evidence that Ras is involved in transcription not only through its binding to Raf, but also through its binding to RafGDS. The activation of RafGDS in human cells, rather than Raf or phosphoinositol-3-kinase (PI3K), plays a central role in the Ras-induced oncogenic transformation. In this paper, we compare the binding of Raf and Raf to Ras, and Raf, another member of the Ras family.

Members of the Ras family share the same core effector binding region (residues 32–40) differing only by a few different binding amino acids. From the solved structures of Ras protein–effector complexes, it is evident that Ras proteins bind to their effectors by establishing an intermolecular β-sheet consisting of two β-strands from the Ras protein and two β-strands from the effector’s Ras binding domain (RBD). All Ras effectors analyzed so far have structurally similar RBDs characterized by a βββββ ubiquitin fold, although the sequence homology in their RBDs is low. Despite the structural similarities of RBDs, Ras proteins bind to different effectors specifically. These findings make the study of selectivity and specificity of Ras proteins intriguing.

The availability of crystal and NMR structures of Ras family members and their mutants as well as structures of Ras–effector complexes, has enabled a number of theoretical studies of the flexibility of proteins, their specificity, and selectivity. Kiel et al. performed an exhaustive mutagenesis analysis of Ras proteins and their effectors by both experimental and theoretical analysis. They made single point interfacial mutants of the Ha–Ras and Rap1A proteins, and their effectors, RafGDS and Raf. Using isothermal titration calorimetry and fluorescence measurements, they determined the thermodynamic binding parameters (binding free energy, enthalpy, and entropy) of Ras proteins to their effectors. Using the FOLD-X energy function, they predicted the binding free energy differences between the mutant complexes. Their assumptions that there are no major conformational changes of the proteins upon a single point mutation or during the complex formation seem to hold up for the majority of the complexes studied since they predicted the binding free energy of different mutants with good accuracy. For Ras–RafGDS complexes, the correlation (R) between calculated and experimental ΔG values is 0.88, with a slope of 0.86; for Ras-Raf complexes, the corresponding values are 0.77 and 0.62. Their study revealed differences in the energy landscape for RafGDS and Raf effectors binding to Ras proteins, and similarity in the importance of different types of interactions in the formation of complexes between Ras proteins and their effectors. According to their results, electrostatic interactions and hydrogen-bond formation are favored over hydrophobic interactions.

Gohlike et al. computed the binding free energies for the Ras-Raf and Ras-RafGDS complexes using the molecular mechanics (MM)-generalized Born surface area (GBSA) approach and 5 ns of MD simulation for the unbound proteins, Ras, Raf and RafGDS, and the complexes, Ras-Raf and Ras-RafGDS. On the basis of snapshots extracted from the simulations, they calculated energy and entropy contributions to the binding free energies. They found that overall the electrostatic interactions disfavor protein–protein binding, not only in the case of Ras and RafGDS proteins, whose net charges are of the same sign, but also in the case of Ras and Raf proteins, which have opposite net charges. On the other hand, Mueggel et al. and Sheinerman and Honig, who both studied the interaction between the RBD of Raf and the Ras homologue Rap1A, found that the electrostatic contribution favors protein–protein association. The main difference between the study of Gohlike et al. and that of the other two groups is in the treatment of the structure of the unbound proteins. Gohlike et al. used the structures of unbound proteins found in the PDB (determined by X-ray diffraction), while the other two groups extracted the individual proteins from the structure of their complex and allowed either no or only a limited relaxation of these.

In the present study, we worked the wild type RafGDS-RBD-Ras-CypNHp, RafGDS-RBD-Rap-CypNHp, and Raf-RBD-Ras-CypNHp complexes, as well as their mutant complexes (55, 21, and 45 complexes, respectively). We used a Poisson–Boltzmann continuum model for electrostatics calculations and performed COMparative Binding Energy (COMBINE) analysis to derive a system-specific quantitative structure activity relationship (QSAR) model for estimating binding free energy differences. COMBINE analysis is based upon the assumption that the binding free energy (ΔG) can be correlated with a subset of suitably weighted energy components (ui) determined from the structures of the two proteins in bound and unbound forms.

\[ \Delta G = \sum w_i u_i + C \]  

The contribution of each interaction energy term is represented by its weight, wi in Eq. (1). The weights are obtained by PLS (partial least squares) analysis using a training set of complexes with experimentally determined binding affinities. The resultant COMBINE model is a system-specific model for binding affinity that, due to the residue-based decomposition, indicates the most important interactions governing binding affinity differences for the complexes studied. The COMBINE model can be used to make predictions of the effects on binding affinity of mutating residues in the complexes. The COMBINE analysis method has proved successful for deriving high quality QSAR models for a variety of protein–ligand complexes including enzyme–inhibitor, enzyme–substrate, protein–peptide, nuclear receptor–DNA complex, and protein–protein complexes. This study is based on the crystal structures of the native enzymes from the Ras family as well as the structures of their mutants in complexes with the effector proteins RafGDS and Raf. It also makes use of thermodynamic parameters.
modynamic measurements for these proteins. First, the structures of the 122 Ras-protein-effector complexes with different interfacial mutations were modeled and energy minimized using the structures of the wild-type protein complexes as templates. The intermolecular energy was decomposed on a residue pair basis and the electrostatic contribution to the desolvation free energy was calculated using the finite difference Poisson–Boltzmann equation. Models that correlate the calculated energy terms with the experimental binding free energy were built by training on only a few complexes (five) selected according to the results of PCA analysis of each set of Ras-effector mutants and used to predict the binding affinity for the other mutants of these proteins. The standard deviations in the errors of predictions (SDEPs, $\sqrt{\sum_{i=1}^{N_S} (\Delta G_{\text{exp}} - \Delta G_{\text{calc}})^2 / N_S}$, for the external data sets of Ras-Raf and Ras-RaLGDS complexes are 1.07 and 1.06, respectively. Such a good predictive performance of the models derived from training on only a few complexes indicates the applicability of the method for planning experiments to study protein binding specificity. The predictive ability of these models is comparable with the predictions achieved with FoldX. The SDEPs of the FoldX-calculated binding free energies for the single mutants of Ras-RaLGDS and Ras-Raf complexes are 0.72 and 0.85, respectively.

METHODS

Molecular Mechanics Modeling

Preparation of mutant complexes

The crystallographic structure of the complex between the Ras binding domain (RBD) of RaLGDS with the E31K mutant of Ras, extracted from the Protein Data bank (code 1LFD), was used for preparing all mutant complexes. For this purpose, E31 in Ras was mutated back to E. There are two complexes (AB and CD) in the asymmetric unit of the crystal cell with the RMSD of their backbone atoms being 0.68 Å. To increase the robustness of the model, we considered both structures in our calculations. The differences between these two complexes are more pronounced for RaLGDS than Ras; the RMSDs between the backbone atoms are 0.60 Å and 0.36 Å, respectively, and the RMSDs of all heavy atoms are 1.25 Å and 0.95 Å, respectively. For building Ras-Ral mutant complexes for which experimentally measured binding affinity is available, mutations were introduced at 11 Ras residues (Q25, V29, E31, D33, E37, D38, S39, Y40, R41, E62, E63) and 12 RaLGDS residues (I14, R16, N23, N25, Y27, K28, S29, K44, D47, K48, H49, D52, E53) (Fig. 1).

The RaRBD-Ras complexes were prepared starting from PDB files 1LFD, containing a Ras-RaLGDS complex, and 1C1Y. 1C1Y and 1GU contain the crystal structures of the complexes of human Rap1a and the RBD of the Ser/Thr protein kinase Raf. According to results obtained with blast (http://www.ncbi.nlm.nih.gov/blast/blast2seq/wblast2.cgi) the sequence identity and sequence similarity between Rap (1C1Y) and Ras (1LFD) RBDs is 57% and 76%, respectively. For Raf and RaLGDS RBDs, sequence alignment with blast/blast2seq is not possible. The alignment with ClustalW (http://www.ebi.ac.uk/ clustalw) for Raf and RaLGDS RBDs gives 18% sequence identity (1A77). We used 1C1Y in our study since Kiel et al. found a better correlation with the experimentally determined $\Delta$G values for the mutants built from 1C1Y than from 1GU. For building mutant Ras-Raf complexes for which experimentally measured binding affinity is available, mutations were introduced at 10 Ras residues (I21, H27, E31, D33, I36, E37, D38, S39, R41, V45) and 8 Raf residues (R59, N64, Q66, R67, T68, V69, K84, V88) (Fig. 2). The Rap–Rap complexes were built using 1GUI, the complex between RapRBD and the Rap K31E, E30D mutant, and 1LFD as templates. Rap from 1GUI and Ras from 1LFD were superimposed, and the RapRBD was replaced by RaLGDS. Modelling was performed in the following way: E31 and D30 in Rap were mutated back to K and E, respectively, and the lowest energy rotamers of K and E were selected. The mutations were introduced at 4 Rap residues (P34, I36, D39, R41) and 12 RaLGDS residues (R16, N23, N25, Y27, K28, S99, K44, D47, K48, H49, D52, E53). The mutations to A, as well as those from Y to F, were performed by simply deleting part of the side chain.

Water molecules included in the calculations were selected using InsightII. The water molecules from the X-ray structures (1LFD and either 1GUI or 1C1Y) close to the interacting proteins (within 6 Å of any of the proteins) were used for further optimization of the complexes. In the Ras-Ras complexes, 189 crystallographic water molecules were retained for the AB subunits and 199 for the CD subunits. The Rap-Ral and Ras-Ras complexes included 89 and 169 water molecules, respectively. The template for the RaLGDS-Ras and RaLGDS-Raf complexes contained 167 Ras residues, 87 RaLGDS residues, GNP bound to Ras and Mg$^{2+}$ coordinated with two GNP oxygen atoms and oxygen atoms of Ras S17 and T35 amino acids. The template for the Ras–Raf complexes contained 167 Ras residues, 77 Raf residues, GNP bound to Ras and Mg$^{2+}$, as well as Ca$^{2+}$ at the surface of the Raf protein stabilized by the side chain oxygen of two glutamates (70 and 71) and the carbonyl oxygen of Gly69.

The nonpolar hydrogens were added using the TLeap module of AMBER8.0. The polar hydrogen atoms of proteins and the bound water molecules were added using the program WHATIF. The assignment of the atom types to the protein and water molecules was done according to the hydrogen bonding environment. Most of the histidines in RaLGDS–Ras and Raf-Raf complexes are on the protein surface far from the binding interface. The closest to the binding interface are His27 in Ras and His49 in RaLGDS which is buried. The ionization of these histidines is determined by local intramolecular contacts and not by whether the protein is free or in the complex.

The complexes with their experimental binding free energies are listed in Tables I and II of Supplementary
Material. The binding free energies of the wild type complexes are $-8.4$ and $-10.0$ kcal/mol,\textsuperscript{26} for the Ras--RalGDS and Ras--Raf complexes, respectively. According to the experimental results, the D47A, D62A, and ESSA mutants of RalGDS (but no mutant of Raf) show higher binding affinities for the wild type Ras. The strongest binding for the Ras--RalGDS complexes was found to be $-9$ kcal/mol between the wild-type RalGDS and Ras: S89A mutant.

PROTEINS: Structure, Function, and Bioinformatics DOI 10.1002/prot
The R41A mutant of Ras binds wild-type Raf (and its N64A mutant) stronger than the wild-type itself.

Energy minimization

The all atom force field of Duan et al. (2003), ff93,49 was used for the proteins and water molecules. Parameters for GNP were derived using ANTECHAMBER and the Mg\(^{2+}\) radius (1.185 Å) was adopted from CHARMM. The parameters were adjusted to reproduce the conformation found in the unbound wild type Ras_GNP structure in the PDB (1CTQ). We found that the relative separation of Mg\(^{2+}\) and GNP in the crystal structure of the Raf/GDP–Ras complex 1LFD is about 1.9 Å larger than in the unbound wild type Ras_GNP structure. In the modeled wild type RafGDS–Ras complex, this difference is less than 1 Å. The TLeap module of AMBER9.0 was used to obtain the topology and coordinate files of each complex. Then, energy minimization of each complex was carried out using the Sandor module of AMBER9.0 in two stages. In the first stage (500 steps), the protein nonhydrogen atoms were restrained to their starting positions by a harmonic potential with a force constant of 32 kcal/(mol Å\(^2\)) while the hydrogen atoms and the water molecules were unrestrained. In the second stage of 200 steps, no constraint was used at all. A nonbonded cutoff of 15.0 Å and a distance-dependent dielectric constant (\(\epsilon = \epsilon_0\)) were used throughout. In each stage, the first 100 steps were performed with the steepest descent algorithm and the rest of the steps were performed with the conjugate gradient method.

During the minimization, the backbone atoms of the mutated proteins did not show observable movement regarding their position in the crystal structure of the proteins, and only water molecules, particularly the additional interfacial water molecules, and some side chain atoms showed significant movements. The RMSD between the optimized and initial structure was never above the RMSD between the two complexes in the asymptotic crystal unit.

Electrostatic Binding Free Energy Calculations

To investigate the electrostatic contributions to the proteins’ desolvation upon complex formation, we performed continuum electrostatic calculations using the program UHBD6.1.52 The electrostatic contribution to the desolvation energy of each protein is defined as the loss of the electrostatic interaction between the solvent and a protein upon binding, as calculated by the two-step procedure described by Perez et al.55 In the first step, the electrostatic interactions within each of the proteins and the surrounding solvent in the absence of the other protein is calculated, and in the second the electrostatic interactions between each of the proteins and the surrounding solvent with the bound pair-protein without partial charges. The electrostatic desolvation energy (\(\Delta G_{\text{desolv,ras}}\) or \(\Delta G_{\text{desolv,Ras}}\)) is given as the difference between the electrostatic energies computed from these two steps. The Poisson–Boltzmann equation was solved using the finite difference method implemented in UHBD6.1. The interior dielectric constant of the protein was set to two and the solvent dielectric constant was set to 78 with an ionic strength of 50 mM and ionic radius of 1.5 Å. The coarse grid spacing was set to 0.60 Å and the fine grid spacing was set to 0.35 Å. The dielectric boundary was defined as the van der Waals surface. Both the coarse grid and the fine grid were dimensioned to 110 \(\times\) 110 \(\times\) 110 with the center on the position of the Co atom of D38/Ras in the case of RafGDS–Rasp complexes, and with the center on the position of the N atom of D37/Ras in the case of Raf–Ras complexes. The two grid centers are 5 Å apart, reflecting the difference in size and position of the two effectors. The coarse grid enclosed the whole protein complex while the fine grid enclosed the interface, including all residues mutated.

Before doing the UHBD calculations, the minimized structures of all the complexes were superposed with the minimized structure of the wild type complex to ensure the same reference coordinates. Then, a separate program was used to convert the superposed structures of the complexes to qed format files for input to UHBD 6.1, with all water molecules removed. In both steps one and two described earlier, the structures of the Ras proteins and their effectors as found in the bound conformation in the complexes were used.

Interaction Energy Decomposition

The energy terms used to define binding free energy (\(\Delta G\)) (Eq. (1)) are the electrostatic desolvation energies of Ras-GppNHp (Ras) or Rap-GppNHp (Rap) and its effector, either Raf-GDP-RBD (RafGDS) or Raf-RBD (Raf), upon binding, \(\Delta G_{\text{desolv,Ras}}\) and \(\Delta G_{\text{desolv,Raf}}\) respectively, and the pair wise electrostatic, \(E_{\text{pair}}\), and van der Waals, \(E_{\text{vdw}}\), interaction energies in energy minimized structures of the complexes. Interactions between each Rasp and each either RafGDS or Rap residue, as well as between GNP and Mg\(^{2+}\) and each effector residue, were determined using the ANAL module of AMBER8.0. A separate code was written to prepare input for GOLPE by extracting the intermolecular electrostatic and van der Waals energy terms, as well as the electrostatic interaction between the partner proteins and their electrostatic desolatation energies calculated with UHBD. For each RafGDS–Ras(p) and Rap–Ras complex, 29,409 \((= 169 \times 87 \times 2 + 3)\) and 26,367 \((= 169 \times 78 \times 2 + 3)\) energy descriptors, respectively, were generated. The binding free energy, \(\Delta G\), was estimated as a weighted linear sum of these energy descriptors as given in Eq. (2).

\[
\Delta G = \Delta G_{\text{desolv,Ras}} + \Delta G_{\text{desolv,Raf}} + \sum_i \Delta G_{\text{pair,ras}}^{\text{prot,ras}} + \sum_i \Delta G_{\text{pair,eff}}^{\text{ras}} + \sum_i \Delta G_{\text{pair,ras}}^{\text{prot,eff}} + C
\] (2)

Chemometric Analysis

The GOLPE4.5 program was used to carry out the chemometric analysis, that is to obtain the weights for

PROTEINS: Structure, Function, and Bioinformatics DOI 10.1002/prot
in Eq. (2), namely the parameters \( w_{\text{denol}}^{\text{Ras}} \), \( w_{\text{renol}}^{\text{Ras}} \), \( w_{\text{redel}}^{\text{Ras}} \), \( w_{\text{redel}}^{\text{Ras}} \), and C. Matrices of the size \( 54 \times 29,409 \), \( 22 \times 29,409 \), and \( 46 \times 29,367 \) were constructed. Each row represents a protein–protein complex described with 29,409 or 29,367 energy terms, the so-called X-variables (Coulombic energy terms, Lennard-Jones energy terms, two desolvation energies, and electrostatic interaction) and the binding free energy (Y-variable).

To reduce the size of the matrix, the X-variables of small absolute value and those showing little variation among the complexes, less than 0.01 kcal/mol, were neglected. To investigate the distribution of the complexes in the energy space defined by these X-variables, a principal component analysis (PCA) was performed. The distances between complexes were measured by the PCA scores. Then, the X-variables were correlated with the Y-variable by PLS analysis to yield initial PLS models of varying dimensionality. To remove the noisy variables and improve the predictive abilities of the PLS models, an X-variable selection procedure consisting of a D-optimal preselection and a fractional factorial design (FFD) was performed. The D-optimal preselection removed nearly half of the X-variables without affecting model quality, and the FFD further removed a few X-variables while retaining uncertain variables. Approximately equal numbers (ca. 10) of the Ras protein residues and its effector residues are included in the interaction energy terms of the final models.

To further evaluate the robustness of the data and the models, we randomly selected three test sets, each containing 41 complexes for training and 12 complexes for external prediction in the case of RasGDS–Ras complexes. In the case of Raf–Ras complexes, we selected three test sets, each containing the same 19 single mutant complexes and 6 different, randomly selected, double mutant complexes for training and the remaining sets of 21 double mutant complexes for external prediction.

RESULTS AND DISCUSSION
Energy Optimized Structures of Mutant Complexes
As in the study of Kiel et al., the investigation is based on the assumptions that single point mutations do not induce major conformational changes in the proteins (Raf, RasGDS, Ras and Rap) and that the conformations of the proteins do not change significantly upon complexation. However, to increase the robustness of the 3D QSAR model, we, instead of using a single structure, used values averaged from two different structures prepared using the two complexes found in the asymmetric unit of the crystal structure (1FLD). The RMSD between the initial and the geometry optimized structure of a mutant was always less than the RMSD between these two complexes (0.7 Å for backbone atoms).

Computed Electrostatic Binding Free Energy Components Correlate With Experimental Binding Free Energies
The calculated \( \Delta G_{\text{binding}}^{\text{Ras}} \) and \( \Delta G_{\text{binding}}^{\text{Ras}} \) values from the Poisson–Rallsmann electrostatics calculations are listed in Table I of Supplementary Material along with the values of the experimental binding free energy, \( \Delta G_{\text{exp}} \). It should be noted that the values reported were computed with a dielectric boundary defined by the protein van der Waals surface. Calculations performed for barnase-barstar complexes revealed that defining the boundary by the solvent accessible surface area leads to greater desolvation costs and a poorer correlation of electrostatic binding free energy with experimental binding affinities.

The linear correlation coefficients between the Ras–RasGDS experimental binding free energy \( \Delta G_{\text{exp}} \) and the electrostatic desolvation energy of Ras \( \Delta G_{\text{desol}}^{\text{Ras}} \) and the desolvation of RasGDS \( \Delta G_{\text{desol}}^{\text{RasGDS}} \) are \(-0.69\) and \(-0.61\), respectively (Fig. 3). For Ras–Raf complexes, a correlation of \(-0.46\) and \(-0.62\) for Ras and Raf, respectively, was obtained, the correlation being weaker in these complexes for Ras. Similarly to the results obtained for barnase-barstar complexes, the unfavorable desolvation energies of both the Ras protein and its effectors are negatively correlated with the experimental binding free energy, i.e., the tightest complexes have the highest desolvation penalties. Apparently, the favorable electrostatic interaction between the partner proteins balances out the unfavorable desolvation effects. The correlation coefficient between the electrostatic interaction energy, calculated by UHBD, and the measured binding free energy is 0.72 for the RasGDS–Ras, and 0.63 for the Raf–Ras complexes. The correlations between the total electrostatic contribution to the binding free energies, calculated by UHBD \( \Delta G_{\text{binding}}^{\text{Ras}} + \Delta G_{\text{binding}}^{\text{RasGDS}} \), and the binding free energies determined by isothermal titration calorimetry, are about 0.67 for both data sets (Fig. 4).

PROTEINS: Structure, Function, and Bioinformatics   DOI 10.1002/prot

Fig. 3. Linear correlations between the RasGDS–Ras experimental binding free energy \( \Delta G_{\text{exp}} \) and the computed continuum electrostatics desolvation energies of Ras (gray dots, thin line) and RasGDS (black triangles, thick line).
Principle Component Analysis for RapGDS–Ras Complexes

The score plot of the first two principal components (PC1 and PC2) for the Rap–Ras complexes is shown in Figure 5. The majority of the complexes is grouped in the lower left quadrant. Exceptions are the complexes with the E37A mutation in Rap which are all grouped in the upper right quadrant and have the largest positive PC1; the complexes with the D33A mutation of Ras, all grouped in the upper left quadrant; and the complexes with the K48A mutation in RapGDS also grouped in the upper left quadrant and having the largest negative PC1 and the largest positive PC2. The complexes with the D38A mutation in Rap are slightly dislocated from the largest group of complexes (grouped in the lower left quadrant) in the direction of the zero value of PC2. The energy terms that dominate in the definition of PCs are desolvation energies and the following RapGDS–Ras electrostatic interactions: K48–D38, K29–D38, Y27–E37, R16–D38, S29–D38, K48–GPN and K48–Mg$^{2+}$. The principal component scores of the complexes in which one or both of these amino acid residues are mutated differ significantly from those in the complexes in which they are not mutated. As a consequence, the particular mutant complexes form groups that are distant from each other and from the rest of the complexes in the score plot (Fig. 5).

Fig. 5. Score plot of the first (PC1) and second (PC2) principal components of the energy terms for 53 RapGDS–Ras complexes. The representative complexes used to build the 5-object COMBINE model are named.

The principal component analysis highlights the energy terms important for the specificity of formation of RapGDS–Ras complexes.

The PC analysis served as a guide for identifying the most representative complexes to include in a minimal representative data set for building a QSAR model by PLS. The correlation between the dominating interactions and the binding affinities will be illustrated by the following PLS analysis.

PLS Analysis for RapGDS–Ras Complexes

The statistical parameters of the final COMBINE PLS models derived for the RapGDS–Ras complexes are given in Table I. The optimal dimensionality, before the variable selection procedure (models not shown in Table I), was determined as two latent variables because the model quality (as measured by fitting (R$^2$ and SDEC) and cross-validation (Q$^2$ and SDEP) parameters) did not increase significantly by adding more latent variables (Fig. 1 of Supplement Material). After the variable selection dimensionality of the model decreased from 2 to 1 LV, see Table I. Exclusion of the complex between RapGDS and RasY40A, WTY40A, resulted in significant improvement of the model performance.

The predicted binding free energies are listed in Table I of Supplement Material. In leave-one-out cross-validation, 90 of the 95 complexes were predicted with an error less than 1.5 kcal/mol. The additional QSAR models were derived using either AMBER, or UHBD energy terms only. Both these models are predictive (Q$^2$ above 0.4). The three 41-complex training sets were randomly selected and PLS models were built. The mean external SDEP values determined for the remaining 12 complexes is 1.02 ± 0.40 kcal/mol (Table I). These SDEP values are mostly close to the SDEP value obtained in leave-one-out cross-validation, indicating robustness of the model.

The complexes with RapGDS-Y27 mutated to Ala were relatively poorly predicted by all these models (with ΔG being about 1−2 kcal/mol, see Table I, supplementary material). Tyr 27 in RapGDS is apparently stabilized by side-chain interactions with Arg 16 in the same protein.

PROTEINS: Structure, Function, and Bioinformatics DOI 10.1002/prot

ID: 40408 Date: 4/2/07 Time: 12:15 Path: J:/Production/PROTVol000006/060260/3B2/C2PROT060250
TABLE I. Predictive Performance of the Final COMBINE Models Derived for Different Sets of Wild-Type and Mutant RaIGDS-Ras Complexes

<table>
<thead>
<tr>
<th>No. of complexes in the training set</th>
<th>X-variables</th>
<th>LV</th>
<th>( R^2 )</th>
<th>SDEC (kcal/mol)</th>
<th>( Q^2 )</th>
<th>SDEP internal for data averaged (kcal/mol)</th>
<th>SDEP external for AB unit only</th>
</tr>
</thead>
<tbody>
<tr>
<td>54</td>
<td>All</td>
<td>1</td>
<td>0.60</td>
<td>0.89</td>
<td>0.54</td>
<td>0.96</td>
<td>--</td>
</tr>
<tr>
<td>53*</td>
<td>All</td>
<td>1</td>
<td>0.66</td>
<td>0.80</td>
<td>0.60</td>
<td>0.88</td>
<td>--</td>
</tr>
<tr>
<td>53</td>
<td>AMBER</td>
<td>2</td>
<td>0.72</td>
<td>0.73</td>
<td>0.62</td>
<td>0.85</td>
<td>--</td>
</tr>
<tr>
<td>53</td>
<td>UHBD</td>
<td></td>
<td>0.56</td>
<td>0.92</td>
<td>0.49</td>
<td>0.99</td>
<td>--</td>
</tr>
<tr>
<td>41# (5x)</td>
<td>All</td>
<td>1</td>
<td>0.68 ± 0.08</td>
<td>0.72 ± 0.14</td>
<td>0.59 ± 0.11</td>
<td>0.81 ± 0.17</td>
<td>1.02 ± 0.40</td>
</tr>
<tr>
<td>5 (PC based)</td>
<td>All</td>
<td>3</td>
<td>0.99</td>
<td>0.02</td>
<td>0.71</td>
<td>1.06</td>
<td>1.13</td>
</tr>
</tbody>
</table>

*The binding free energies predicted by this model are given in the last column of Table I of Supplement Material.

*The mean values of three models derived for the three randomly selected 12-object test data sets are given.

(transition between the hydroxyl O and NH₂ is less than 3 Å). The mutation of Y27 to Ala abolishes this stabilization, so this might influence the protein stability, and accordingly reduce the binding affinity. Since our COMBINE models do not consider the change in protein stability; the predicted binding affinities are more favorable than the measured ones.

To examine the applicability of the COMBINE procedure for planning efficient experiments to alter binding by mutation, we prepared a 3D QSAR model using only five complexes selected on the basis of the PC analysis score plot that we selected one object from each of the five distinct groups. The following objects were selected: K48AD38A, WTD33A, S298ES37A, S298AD38A, H498AS38A. In this way, the training set spanned the majority of the PC space. The model derived with these five objects has good predictive performance: before FFD, the \( Q^2 \) value is 0.52, and SDEP for the other 48 (test) complexes is 1.17, and after FFD the \( Q^2 \) value is 0.71, and SDEP for the other 48 complexes 1.13, (Table I). The external prediction quality of this model is similar to the external prediction quality for the models derived using a training set of 41 randomly selected complexes.

To investigate which of the energy terms were highly weighted in the COMBINE model, we listed the thirty weights with the highest coefficients. The RaIGDS amino acid residues that appear in these terms are I15, R16, D22, N23, G24, N25, M26, Y27, K28, S29, I30, R43, K44, A45, D47, K48, H49, N50, D52, E53, and D60. The Ras amino acid residues that appear in these terms are K5, K16, N25, E31, D33, P34, T35, I36, E37, D38, S39, Y40, R41, D44, L54, L65, D67, N61, E63, R68, and GNP and the Mgp2 ion. The list includes all the RaIGDS and Ras hot-spot residues given by Kiel et al.\(^2\) The top 10 PLS coefficients of the selected X-variables (energy terms), namely the weight parameters \( w_1 \) and \( w_2 \) and \( w_3 \) and \( w_4 \) in Eq. (1), are plotted in Fig. 6.

By multiplying these weights with the corresponding energies, we obtain the contribution of the energy terms to the computed binding affinity. However, some of these terms make contributions that are opposite to their expected contribution. For example, increasing the protein electrostatic desolvation energy leads to more favorable computed binding affinity. This is because increased electrostatic desolvation energy correlates with increased favorable electrostatic interactions (see previous section). The model derived using AMBER energy terms is the most appropriate to investigate the influence of the certain pairwise intermolecular amino acid residue interactions on the binding free energy. For this purpose we multiplied weights above 0.01 with the corresponding energies calculated for the complex between wild type proteins. The most stabilizing interactions are those between Ras:K38 and Ras residues K28 and K48 (-1.63, and -0.85 kcal/mol, respectively) and the Ras:Y27-Ras:Y27 interaction (-1.46 kcal/mol, Fig. 7). Interactions with unfavorable contributions to the binding affinity make contributions with much lower magnitude. The highest are Ras:E31-Ras:D52, Ras:E31-Ras:D47, and Ras:S39-Ras:Y27 interactions, all below 0.07 kcal/mol.

Beside models for RaIGDS–Ras complexes, we also derived a model with a training set of 22 RaIGDS-Rap complexes. The model based on UHBD desolvation energies and AMBER energy interaction terms has \( R^2 \) and \( Q^2 \) values of 0.62 and 0.48, respectively, and SDEP and SDEP values of 0.72 and 0.85 kcal/mol, respectively. The COMBINE model, even when trained with a small number of carefully selected mutants, predicts the RaIGDS–Ras binding affinity fairly well. However, it should be
born in mind that the COMBINE model is system-specific and the good prediction of the Ras–Raf/GDS binding affinity is not a guarantee that the binding affinities between the other proteins from the Ras family and their effectors (for example Ras/GDS–Rap and Raf–Ras) will be correctly predicted. Robustness of the models derived for the Ras/GDS–Ras complexes (Table 1) was tested in a way that they were used to predict binding affinities for Ras/GDS–Rap complexes. The SDEPs determined for the 22 Ras/GDS–Rap complexes were between two and three. The best prediction of the Ras/GDS–Rap binding affinities was obtained with the model trained with five objects selected on the basis of PC analysis (SDEP = 2.24), followed by predictions with the UHBD and AMBER models trained with the 53-object data set. After correction for the difference in average energy of the datasets, \( \sqrt{\sum (\Delta G_{\text{test}} - \Delta G_{\text{train}})^2/n} \), where \( \Delta G_{\text{test}} \) and \( \Delta G_{\text{train}} \) are the mean free binding energies of the test (Ras/GDS–Rap) and the training (Ras–Raf/GDS) sets, SDEP improves. The SDEPs determined for the test set of 22 Ras/GDS–Rap complexes with the 5-object (PC based), UHBD and AMBER models (Table 1) are 1.26, 1.24, and 1.21 kcal/mol, respectively.

**PCA Analysis for Ras–Raf Complexes**

In the score plot of the first two principal components (PC1 and PC2) for the Ras–Raf complexes (Fig. 2 of Supplementary Materials), the majority of the complexes, as in the PC analysis for Ras–Raf/GDS, is grouped in the lower left quadrant. Exceptions are the complexes with the E37A mutation in Ras which are all grouped in the upper quadrant and have the largest positive PC2; one of them, the K84AE37A complex, is positioned alone in the upper right quadrant. The complexes with either the RafK84A or RasE31A mutation are grouped in the lower right quadrant and have the largest negative PC2 and the largest positive PC1.

The energy terms that dominate in the definition of PCs are the electrostatic interactions between Ras–E37 and Raf residues R59, N64 and T68, those between Ras–D38 and Raf residues T68 and K84, and the GNP–RafK84 interaction. In the K84AE37 (K84AE37A) mutant, most of these interactions are abolished, as it has an isolated position in the loading plot. This is in agreement with the experimental results according to which the binding free energy for this mutant is 3.5 kcal/mol less favorable than the binding free energy for the wild type protein complex (−6.5 vs −10 kcal/mol).

**PLS Analysis for Raf–Ras Complexes**

3D QSAR models were derived for a training set of 46 Raf–Ras complexes built using the coordinates of the complex in PDB file 1CLX to model the Raf mutants and the B subunit from the PDB file 1LFD to model the Ras mutants. The models are of good predictive performance, see Table II. COMBINE models determined for three 25-object training sets, randomly extracted from the 46-object data set, were used to predict the binding affinities of the complexes in the remaining 21-object data sets. The mean external SDEP calculated from these three models is 1.11 ± 0.07 kcal/mol (the experimental binding free energies span about 5 kcal/mol, from −10.6 to −5.4). The complexes with the I30A Ras mutation are poorly predicted by all models, with R59G63 being an outlier. To examine the influence of slight variation in geometry on the COMBINE model performance, two additional 3D QSAR models were derived: (a) for the set of further optimized complexes, and (b) for the set of data averaged over the complexes obtained using Ras from two different asymmetric units of 1LFD (B and D).

![Diagram](image)

**Fig. 7.** The favorable contributions (< −0.05 kcal/mol) to the binding free energy of the wild type Ras/GDS–Ras complexes (AMBER model, see Table I). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

**Table II. Predictive Performance of the COMBINE Models Derived for Raf–Ras Complexes**

<table>
<thead>
<tr>
<th>Number of complexes</th>
<th>X-variables</th>
<th>LV</th>
<th>R²</th>
<th>SDEC (kcal/mol)</th>
<th>Q²</th>
<th>SDEP Internal (kcal/mol)</th>
<th>SDEP_AB external* (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>46</td>
<td>All</td>
<td>5</td>
<td>0.85</td>
<td>0.48</td>
<td>0.69</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>AMBER</td>
<td>5</td>
<td>0.81</td>
<td>0.34</td>
<td>0.64</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>UHBD</td>
<td>3</td>
<td>0.48</td>
<td>0.80</td>
<td>0.40</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>5 PCA based</td>
<td>All</td>
<td>2</td>
<td>0.96</td>
<td>0.26</td>
<td>0.73</td>
<td>0.66</td>
<td>1.20 (41)</td>
</tr>
</tbody>
</table>

*Number of objects in the test data set is given in brackets.

**PROTEINS: Structure, Function, and Bioinformatics** DOI 10.1002/prot

Id: 40408 Date: 4/2/07 Time: 12:16 Path: J:\Production\PROTVol090000\060260\062C2\J2PRQ7060260
The averaging did not significantly influence the predictive performance of the model, while resampling slightly reduced the model performance (data given in Table III of Supplementary Materials).

The models derived using either only the UHBD or only AMBER energy terms are also listed in Table II. Both are predictive, but the latter has better performance. Unlike the models derived for the RaIGDS–Ras complexes, the ‘AMBER-only’ model has slightly lower predictive performance than the model obtained using both AMBER and UHBD variables.

To determine which of the highly weighted pairwise, intermolecular energy terms are the most important for the free energy in a strongly binding complex, we multiplied weights in the COMBINE model ($R^2 = 0.85, Q^2 = 0.69$) with the corresponding energy terms calculated by AMBER for the wild type Ras–Raf complex. The greatest contributions to the binding free energy are shown in Figure 8. It is interesting that the Van der Waals interaction plays a more significant role in this model than in the model derived for the RaIGDS–Ras complexes. The highest Van der Waals interaction energy term favorably contributing to the binding free energy is that between RasN64 and RafN89. Beside the amino acid residues that were mutated in the study, a few nonmutated sites, namely RasY40 and Raf residues T57 and R89, appear as important. Ras residue Y40 strongly interacts with RafQ66 (E_{AMBER} = −3.2 kcal/mol) and in this way stabilizes the complex ($ΔG_{ Ras} = ca. −0.6 kcal/mol$). The contribution of the unfavorable pairwise energy terms is higher in the Ras–Raf COMBINE models than in the Ras–RaIGDS models. The most destabilizing energy terms are the electrostatic interactions between RasS39-RafN88 and RasR41-RafN64 with 1.4 and 0.7 kcal/mol, respectively. Both these interaction are attractive, but because of negative weights unfavorably contribute to the predicted binding free energy. Possible explanation is that desolvation cost of Raf residues N64 and R89 is not compensated with the attractive electrostatic interactions with Ras.

To improve the favorable RasS36-RafS57 interaction, we computationally mutated RafT57 to tyrosine, resulting in T57Y being expected to bind as well or slightly better than WT; according to the COMBINE model $ΔG$ is $−9.2$ and $−9.4$ kcal/mol for the wild type proteins and T57YWT, respectively. Besides the RasT36-RafT57 interaction ($ca. −0.15$ kcal/mol), some other favorable contributions to the binding free energy are computed to improve slightly upon this mutation.

The COMBINE Models for Cross Prediction of Binding Affinities for RaIGDS–Ras and Raf–Ras Complexes (the Reduced COMBINE Models)

The number of X variables, as defined in Eq. (2), in the COMBINE models for RaIGDS–Ras(p) and Raf–Ras complexes differ because of the different sizes of the effector proteins, and a COMBINE model derived for the one set of these Ras-effector complexes cannot be used directly to predict binding affinity at the other. However, the reduced, UHBD model consisting only of the proteins desolvation energy ($ΔG_{ desolv}$, $ΔG_{ Raf}$), the first two terms in Eq. (2), and the electrostatic interaction between them, $E_6$, can be used for the ‘cross prediction’ between the sets of different protein complexes. Using the ‘UHBD3’ model (Table II) derived for the Ras–Raf set of complexes, we predicted binding affinities for the RaIGDS–Ras and RaIGDS–Raf mutants. The SDEPs obtained with the model that accounts for differences in the average binding free energy of the Ras–Raf and Ras–RaIGDS complexes are 1.20 and 1.15 kcal/mol for Ras–RaIGDS and Raf–Raf complexes, respectively. Beside this, the models including the interaction energy terms were also adjusted to be used for cross validation of two different sets of complexes: RaIGDS–Ras(p) and Raf–Ras. For this purpose, the whole effector was treated as one group interacting with each amino acid residue of Ras protein. In this way, we derived two different COMBINE models, one based on AMBER Van der Waals and electrostatic interaction energies, and the other based on UHBD electrostatic interaction energies. Beside the Ras-residue based interaction energies, the desolvation energy terms and the total PB electrostatic interaction energy were also included. The number of X-variables used to derive these models is 341 (169 × 2 + 3), and 171 (169 + 2), respectively. The UHBD_171 model is more robust than AMBER_341 and has better external predictive performance (Table III). For 39 out of 53 RaIGDS–Ras complexes, the binding affinity is predicted by the UHBD_171 model with an error less than 1.5 kcal/mol (Fig. 9). The correlation between experimental and predicted binding affinities for RaIGDS–Ras complexes is about 0.61. Binding affinities for complexes with RaIGDS-Y40 mutated to either alanine or phenylalanine.
and RasGDS-RasY27A are poorly predicted, but these complexes were poorly predicted even with the COMBINE model derived for the RasGDS–Ras complexes. The electrostatic interactions between E37, D38, S39, Y40, D57 and GNP and Raf are the largest favorable contributions to the predicted binding affinity.

Recently Kiel et al. made the M26K, D47K, E1/D34K RasGDS mutant and found that it binds to Ras 14-fold faster and 25-fold tighter compared with the wild-type. These three amino acid residues appear as interaction partners in the thirty highly weighted terms of the COMBINE model derived for the RasGDS–Ras complexes. However, their contribution to the binding free energy is much weaker than those of the RasGDS residues K28 and K48. The binding affinity for this complex predicted by the reduced UHBD based model is almost identical to the binding affinity predicted for the wild type RasGDS, that is, $-9.09$ vs $-9.07$ kcal/mol. Apparently the COMBINE model predicted tight binding of the mutant, but not the large gain in binding affinity observed experimentally.

Possible Small Molecule Inhibitor Binding Sites

Trosset et al. showed that it is possible to inhibit a protein–protein interaction by binding of a small molecule that tightly interacts with the hot spots at the protein interface. Such a hot spot position for the Ras would be in the vicinity of E37 and D38. However, a small molecule could be utilized to specifically inhibit binding of only one effector. According to the results of COMBINE analysis, we expect that a small molecule that would bind into the clefts between E31 and D33 and/or D33 and T35 should more significantly decrease the binding affinity of Ras to RasGDS than to Raf. Namely, the side chain oxygen of D33, E31 and T35, as well as the backbone oxygen of P34, tightly interact with Ras residues N50 and lysines 28 and 48 (highly weighted interactions in the COMBINE models derived for RasGDS–Ras complexes), while there are no such close interactions between Raf and these Ras residues (Raf is at this part of the Ras surface more distant from it than Ras). Another possible place for binding a small molecule with discriminating influence to Ras–Raf and Ras–RasGDS complexes might be in the vicinity of Y40, R41, and D54. At this part of the Ras surface, Raf is closer to it than Ras, and its residues N64, Q66, and R67 favorably interact with Ras residues Y40, R41, and D54.

**Concluding Remarks**

The electrostatic contributions to binding computed by solving the finite difference Poisson–Boltzmann equation, with a van der Waals surface boundary definition, favor formation of complexes between Ras proteins (Ras and Rap) and their effectors (Raf and RasGDS).

The 3D QSAR models derived using COMBINE analysis are predictive not only for the complexes of the same Ras protein and its effector, i.e. for the mutants of the same partner proteins, but also for the sets of different Ras-effector protein complexes. The binding affinities of complexes with 136A and Y40A(F) Ras mutants were poorly predicted by all models. There are two possible explanations for this, one is changed (increased) stability of the mutant and the other is conformational change that improves complex formation. These effects could not be detected with the standard COMBINE approach.

The Ras amino acid residues that all, AMBER and UHBD based, COMBINE models for both datasets, RasGDS–Ras and Rap–Ras complexes, revealed as significantly stabilizing are E37, D38, and D57. Besides, in both data sets, an important stabilization is achieved by interaction between GNP and the effector protein. The other residues of the Ras protein that appear important for the models are K5, K16, E31, D33, P94, I56, S39, Y40, R41, and D54. The Raf–Ras complexes are specifically stabilized by the electrostatic interactions: Ras: Y40–Raf–Q66, Ras:S39–Raf–Q66, Ras:S39–Raf–R67, Ras: R41–Raf–Q66, and Ras:D54–Raf–R67, and the RasGDS–Ras complexes complexes are specifically stabilized by

**PROTEINS: Structure, Function, and Bioinformatics** DOI 10.1002/prot
the electrostatic interactions between Ras residues E31 and P34 and Raf residue K48.

The present analysis enabled us to propose Ras amino acid residues important for discriminating binding of Raf and RafGDS. The results suggest that a small molecule that would specifically bind into the cleft between E31 and D33 and/or D31 and T35 could decrease the binding affinity of Ras to RafGDS more than to Raf. On the other hand a molecule bound in the vicinity of Y40, R41, and D54 could decrease the binding affinity of Ras to Raf more than to RafGDS.

The COMBINE models derived for only a few complexes selected according to the results of PC analysis successfully predicted binding affinities for the remaining, external set of complexes. This result allows us to point to the feasibility of this method for aiding experimental design. A number of in silico mutants accompanied by a PC analysis could be used as an initial base for selecting representative mutants to be analyzed experimentally. From the calculated energy interaction terms and the experimentally determined binding affinities for the selected mutants, the PL5 model could be derived. Finally, binding affinities for the other, in silico prepared mutants would be predicted and those with the highest binding affinities would be selected for further experimental study. However, an important assumption of such an approach is that the mutations do not induce significant conformational changes in the protein.

The electrostatic components of $\Delta G$ calculated with UHBD correlate slightly better with the measured value than $\Delta G$ computed with the COMBINE model with the UHBD terms only (0.67 vs 0.61). However, $\Delta G$ values are by UHBD significantly overestimated, on average about 6 and 12 kcal/mol for RafGDS-Ras and Raf-Ras complexes datasets, respectively. On the other hand, binding affinities for 77% of the analyzed complexes are predicted by the COMBINE model with an error less than 1.6 kcal/mol. To estimate the relative binding affinity for a set of protein complexes standard Poisson-Boltzmann electrostatic free energies can be computed, and this can be done in the absence of any experimental data on binding affinities. On the other hand, if experimental data are available for at least a few representatives, the COMBINE method may be useful. These two methods nicely complement one another since a reliable 3D structure of the complex is needed for both methods, and once the UHBD energies have been calculated, a PC and PL5 (if affinity measurements are available) analysis can easily be performed and valuable information obtained.

If we compare COMBINE methods with more general methods to estimate binding affinity we can give arguments pro and contra. On the one hand, being system specific, it may capture features of the system important for binding affinity that are difficult to detect with a general scoring function. On the other hand, a drawback is that, being trained for a specific system, the accuracy of predictions falls off as complexes in the test set become more different from those in the training set. An advantage of the COMBINE method in comparison with most general scoring functions or methods to estimate binding affinity is that it explicitly suggests which amino acid residues should be targeted to adjust the binding affinity, either by mutation or by binding of a small molecule.

ACKNOWLEDGMENTS

We thank Prof. C. Herrmann for providing us data and for valuable discussion.

REFERENCES


PROTEINS: Structure, Function, and Bioinformatics DOI 10.1002/prot
26. Haang L, Weng X, Hofer F, Martin OS, Kim SH. Three-dimen-
sional structure of the Ras-interacting domain of RaGDS. Nat
27. Geyer M, Hermann C, Wohlgemuth S, Wittinghofer A, Kalbit-
zer HR. Structure of the Ras-binding domain of RaGEP and
implications for Ras binding and signaling. Nat Struct Biol
28. Walker EH, Periasc O, Bied C, Stephens L, Williams RL. Struc-
tural insights into the phosphoseryl 5-kinase catalysis and signi-
29. Steiner G, Kremer W, Linnemann T, Herrmann C, Geyer M,
Kalbitzer HR. Sequence-specific resource assignment of the
30. Grounev W, Huber F, Grunewald P, Spornier M, Wohlgemuth S,
Herrmann C, Kalbitzer HR. Solution structure of the Ras bind-
ing domain of the protein kinase Bz2 from Schwannocarya
31. Scheifele R, Grunewald P, Wohlgemuth S, Kabach W, Th,
Wigler M, Wittinghofer A, Herrmann C. The Ras-Bz2 complex:
structural basis for Ras effector recognition in yeast. Struc-
32. Golikhe R, Hahn LA, Case DA. Change in protein flexibility
upon complex formation: analysis of Ras-Raf using molecular
33. Golikhe R, Case DA. Converging free energy estimates: MM-
PB.GBSA studies on the protein-protein complex Ras-Raf. J
34. Golikhe R, Kiel C, Case DA. Insights into protein-protein bind-
ing by binding free energy calculation and free energy decompo-
sition for the Ras-Raf and Ras-RaGDS complexes. J Mol Biol
35. Muller J, Zweis T, Karnesel A. Electrostatic contributions to
protein-protein binding affinity: application to Rap/Raf interac-
36. Shinamman RB, Honig B. On the role of electrostatic interac-
37. Kiel C, Wohlgemuth S, Rouassou F, Schwenkowt J, Pierking-
off J, Wittinghofer A, F, Serrano L. Recognizing and defining
true ras binding domains; ii: in silico prediction based on homol-
38. Madura JD, Briggs JM, Wade RC, Davis ME, Luty BA, Iln A,
binding affinities by comparative binding energy analysis. J
structurally optimized Ras-binding f2 id sequence alignment
42. Wade RC. Derivation of QSA for 3 structural models of
list protein ligands using COMBINE analysis. In: Holte
HD, Stipp S, editors. Rational approaches to drug design: 13th
European symposium on quantitative structure-activity relation-
43. Pastor M, Gago F, Cruciani G. Comparative binding energy
COMBINE analysis on a series of glycerol phosphate inhibitors: comparison with GRID/GOLPE methods. In: Gunder-
tofl K, Jorgensen FS, editors. Molecular modeling and predic-
44. Wang T, Wade RC. Comparative binding energy (COMBINE)
analysis of influenza neuraminidase inhibitor complexes. J
45. Lorenzo JF, Pastor M, Cruciani G, Galet K, Centeno NB, Gago
F, Saznik F. 3D-Q SAR methods on the basis of ligand receptor
complexes. Application of COMBINE and GRID/GOLPE methods
and analysis of CYP1A2 ligands. J Comput Aided Mol Des
46. Tomk S, Kajic-Prodic B. A quantitative model for predicting
energy of complex formation (CMEF) and burden of the b cell
inhibitory 24-3 regulatory sequence of the hla-e gene. J Mol Graph
47. Tomk S, Kajic-Prodic B, Koluvay I. Stereospecificity of
the human b2 adrenoceptor 3a2p-adenosine receptor. J Med Chem
2002;45:4828–4837.
48. Tomk S, Nilsson L, Wade RC. Nuclear receptor-DNA binding
49. Tomk S, Wade RC. COMBINE analysis of opae-peptide complexes to relate structure to binding
50. Tomk S, Nilsson L, Wade RC. Nuclear receptor-DNA binding
51. Tomk S, Wade RC. COMBINE analysis of nuclear receptor-DNA
binding specificity: comparison of two datasets. Chemic Acta
52. Wang T, Tomk S, Gobobinov R, Wade RC. How optimal are
the binding energies of b2 and b3a1 receptors? Biophys J 2004;85:
1563–1574.
proteins based on condensed-phase quantum mechanical calcu-
55. Vriend G. WHAT IF: molecular modeling and drug design
56. Houw RW, Sander C, Vriend G. Positioning hydrogen atoms by
optimizing hydrogen bond networks in protein structures. Pro-
57. Madura JD, Briggs JM, Wade RC, Davis ME, Luty BA, Ilin A,
58. Perez C, Pastor M, Ortiz AR, Gago F. Comparative binding
energy analysis of HIV-1 protease inhibitors: incorporation of
solvent effects and validation as a powerful tool in receptor-
59. Baroni M, Costantino G, Cruciani G, Rignani D, Valigi R, Cem-
eni S. Generating optimal linear PLS estimations (GOLPE): an
advanced chemicometric tool for handling 3D-Q SAR problems.
60. Kiel C, Selzer T, Shaul Y, Schwerin G, Herrmann C. Electro-
statically optimized Ras-binding f2 id sequence alignment
61. Tresset J-Y, Dalvit C, Knapp S, Faouli M, Veronese M, Mantze-
interactions: the discovery of druglike b-catenin inhibitors by combi-
ning virtual and biophysical screening biophysical screening.
PROTEINS: Structure, Function, and Bioinformatics DOI:10.1002/prot

ID: 40408
Date: 4/2/07
Time: 12:16
Path: J:/Production/PROT/Vol00006/06/2620/3B2C2PROT060260

J.D.: ZTE
Customer ID: 21321
Caldmus Art: PROT21321
Date: 4-02-FEBRUARY
Stage: I
Page: 13

BINDING ENERGETICS OF RaGDS MUTANTS